شبكة الكلمات العربيَّة

«المنهجيَّة والتَّطبيق»

د. المُعتَزّ باللهَّ السَّعيد طه(*)

الكلمات الدَّالَّة

شبكة الكلمات، الأنطُولوجيا، المُدَوَّنة اللُّغَوِيَّة، صناعة المُعجم.

ملخص

تَعرِضُ هذه الورقةُ مدخلاً إلى الشَّبكات اللُّغَوِيَّة للكلمات WordNets باعتبارِها تمثيلاً معرفِيًّا واسِعًا لجوانب اللُّغة الإنسانِيَّة. ويُقَدِّمُ الباحثُ – فيها – منهجًا مُقتَرَحًا لبناء شبكةٍ للكلمات العَرَبِيَّة Arabic Word Net انطلاقًا من الشَّبكات المِعيارِيَّة المُنجَزة، مع مُراعاة طبيعة اللُّغة العربيَّة الاشتقاقيَّة ونظامِها المُعجَمِيّ والدَّلاليّ – على مُستوى المباني والمعاني. كما يعرِضُ الباحثُ لجوانب الإفادة من شبكة الكلمات المنشودة في تطوير الصِّناعة المُعجَمِيَّة للعربيَّة المُعاصِرة، مع العناية بالمُعجَمات المُتَخَصِّصة.

(*) باحث في علم اللغة بكلية دار العلوم بجامعة القاهرة.

١- مقدمة

١.١- شبكة الكلمات.

يُمكنُ تعريفُ شبكة الكلماتِ بأنَّها قاعدةُ بياناتٍ مُعجَمِيَّةٌ Lexical Database؛ أو بعبارةٍ أخرى: شبكةٌ من العلاقات المُعجَمِيَّة Lexical Relations، تُعالَجُ فيها نُصُوصُ اللُّغة بتعيين أقسامِ الكلام Parts of Speech (POS)، ثُمَّ تُصَنَّفُ في مجموعاتٍ مُترابطةٍ من المُترادِفات Synonyms / Sets of Cognitive Synonyms (Synsets) الَّتي تُعَبِّرُ عن مَفاهِيمَ ثابتةٍ[١]، وتتَّفقُ في معانيها الدَّلاليَّة – وإن اختَلَفَت في المعاني السِّياقِيَّة؛ وتَرتَبِطُ هذه المُترادفات فيما بينها بشبكةٍ من العلاقات الدَّلالِيَّة Semantic Relations.

٢.١- الخلفيَّة والإرهاصات.

١.٢.١- شبكة كلمات برينستون PWN.

ظَهَرَت فِكرَةُ شَبَكة الكلمات في اللُّغَة الإنكليزِيَّة لتكونَ أداةً لتطوِير صناعة المُعجَم Lexicography وتطبيقات الذَّكاء الصُّنعيّ Artificial Intelligence (AI) ومُعالَجَة اللُّغات الطَّبِيعِيَّة Natural Language Processing (NLP). وكانت البدايةُ الحقيقيَّة لشبكات الكلماتِ في عام ١٩٨٥م، إذ بدَأَ العَمَلُ في بِناء وتطوير شَبكةٍ كامِلَةٍ لِمُفرَدات اللُّغَة الإنكليزِيَّة في مُختَبَر العُلُوم الإدراكِيَّة بجامعة برينستون Princeton الأمريكيَّة[٢]؛ وتَمَكَّنَ الفَرِيقُ الَّذي أشرَفَ عَلَيه عالِمُ النَّفس الأمريكِيّ جورج ميلر George Miller مِن بِناء الشَّبَكَةِ وتَعبِئَتِها يَدَوِيًّا قبلَ إتاحَتِها على الشَّبَكة العنكبوتيَّة

(1) Fellbaum, C. (1998). WordNet: an electronic lexical database. Cambridge, Mass: MIT Press.

(2) Fellbaum, C., (1998). Ibid.

تحت مُسَمَّى «شبكة كلمات برينستون Princeton WordNet». وقد بلغَ عددُ الكلمات المُفرَدة Unique words في هذه الشَّبكة 155287 كلمة، تَضُمُّ الأقسامَ الكلاميَّة: الاسم Noun = 117798 كلمة، والفِعل Verb = 11529 كلمة، والصِّفة Adjective = 21479 كلمة، والظَّرف (الحال) Adverb = 4481 كلمة (PWN 3.0 - 2011).

ويبلُغ عدد المجموعات التَّرادُفِيَّة Synsets الَّتي تضُمُّ كلمات الشَّبكة 117659 مجموعة (82115 للأسماء، 13767 للأفعال، 18156 للصِّفات، 3621 للظُّرُوف)، حيث تترابطُ كلمات كُلّ مجموعة عن طريق شبكةٍ مُوَسَّعةٍ من العلاقات الدَّلاليَّة، نستطيعُ التَّمثِيلَ لها بالعلاقات المُوَضَّحة في الجدول التَّالي[٣]:

العلاقة	المفهوم (التَّعبير الرِّياضيّ)	المثال
الاشتماليَّة Hypernyms	**Y** is a hypernym of **X** if every **X** is a (kind of) **Y**	**canine** is a hypernym of **dog**
النَّوعِيَّة Hyponyms	**Y** is a hyponym of **X** if every **Y** is a (kind of) **X**	**dog** is a hyponym of **canine**
الكُلِّيَّة Holonym	**Y** is a holonym of **X** if **X** is a part of **Y**	**building** is a holonym of **window**
الجُزئيَّة Meronym	**Y** is a meronym of **X** if **Y** is a part of **X**	**window** is a meronym of **building**
التَّلازُمِيَّة Troponym (خاصَّة بالأفعال)	the verb **Y** is a troponym of the verb **X** if the activity **Y** is doing **X** in some manner	to **lisp** is a troponym of to **talk**
الاستلزاميَّة Entailment	the verb **Y** is entailed by **X** if by doing **X** you must be doing **Y**	to **sleep** is entailed by to **snore**

الجدول 1: من العلاقات الدَّلاليَّة في شبكة كلمات برينستون PWN

(3) Fellbaum, C., (1998). Ibid.



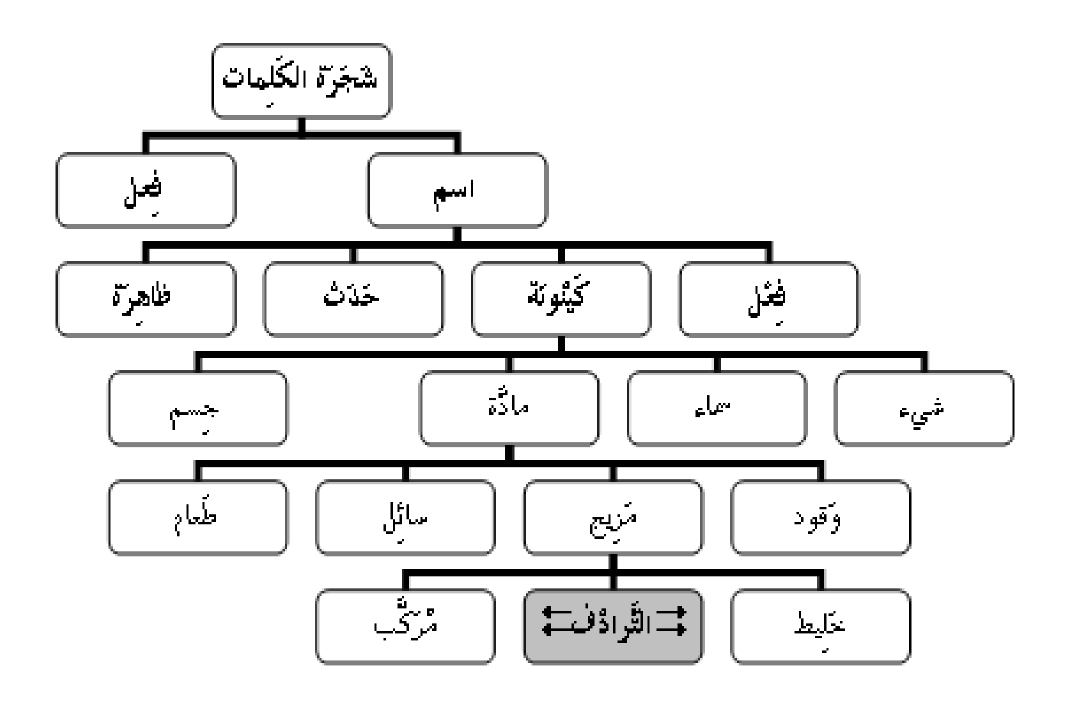
الشَّكل ١: نموذج من شَبَكَة كلمات برينستون للُّغة الإنكليزِيَّة «PWN 3.1 - 2011»

١.٢.٢- شبكة الكلمات العربيَّة AWN.

في المُدَّة من ٢٠٠٥م إلى ٢٠٠٧م، سعى فريقٌ بحثيٌّ – في جامعات برينستون وبرشلونة ومانشِستر – لِبناءِ شَبكةٍ للكلمات العَرَبِيَّة على غرار شبكة برينستون. وأشرَفَت الباحثةُ الألمانِيَّة كريستيان فيلباوم Christiane Fellbaum – الَّتي شارَكَت من قبل في بناء شَبَكَة كلِمات برينستون بقِيادَة جورج ميلر – على بناء شبكة الكلمات العربيَّة AWN؛ إلاَّ أنَّ العملَ تَوَقَّفَ نتيجةَ نقصِ الموارد المُتاحة لفريق العمَل.

وَتَقُومُ الفِكرَةُ الأساسِيَّةُ لمشروع شَبَكة الكَلِمات العَرَبِيَّة AWN على تزويد قاعدة بيانات الشَّبكة بمجموعةٍ من المُفرَدات «المفاهيم» المُتَرجمةِ عن شبكة برينستون

انطلاقًا من نَظَرِيَّة الحُقُول الدَّلاليَّة Semantic fields حيثُ تتشكَّلُ الشَّبكةُ في بِناء شجَرِيٍّ تُصَنَّفُ فيه المُفرَداتُ إلى أسماء وأفعال، ويَتَفَرَّعُ عن كُلٍّ منهما مجموعةٌ مِن الحُقُول الَّتي تَضُمُّ عددًا مِن المُفرَدات المُوَزَّعة على حُقُولٍ فرعيَّةٍ، وترتبطُ هذه المُفرَداتُ فيما بينها بعلاقاتٍ دَلاليَّة(٤). ويبلُغُ عددُ كلماتِ شبكة الكلمات العربيَّة AWN 17561 كلمة مُفرَدة؛ أمَّا عدد المجموعات التَّرادُفِيَّة Synsets الَّتي تضُمُّ كلمات الشَّبكة فيبلُغ ٧٨٢٢ مجموعة.



الشَّكل ٢: نموذج من شَجَرَة الكَلِمات العَرَبِيَّة في شَبَكَة الكلمات العربيَّة
«AWN 2.0 – 2007»

(4) Black, W., Elkateb, S., Rodriguez, H, Alkhalifa, M., Vossen, P., Pease, A. and Fellbaum, C., (2006). Introducing the Arabic WordNet Project, in Proceedings of the Third International WordNet Conference, Sojka, Choi, Fellbaum and Vossen eds.

١.٣- شبكة الكلمات بينَ الشَّبكة الدَّلاليَّة والأنطولوجيا.

تتكوَّن شبكةُ الكلمات من مجموعةٍ من الشَّبكات الدَّلاليَّة Semantic Nets الَّتي تربطُ بينَ المفاهيم باعتبارِها شكلاً من أشكال التَّمثيل المَعرِفيّ؛ وفي الوقت ذاتِه يتكوَّن من الارتباط بين مجموعةٍ من شبكات الكلمات هيكلٌ مَعرِفيٌّ ضخم، يُستَعمل في تطبيقاتٍ عديدةٍ تتبعُ الذَّكاءَ الصُّنعيَّ وعلومَ المكتَبات وصناعة المُعجَمات وتطبيقات الويب الدَّلاليّ Semantic web، ويُطلَقُ عليه «الأنطولوجيا Ontology». وتُمَثِّلُ الأنطولوجيا مجموعةً من الكُتَل المَعرِفيَّة الَّتي تتلاحمُ عناصِرُها لتُكَوِّنَ تمثيلاً لمجموعاتٍ مُنَظَّمة من المفاهيم المُشتركة في حُقُول دَلاليَّة مُحدَّدة، مُترابطة بعلاقات دَلاليَّةٍ(٥)، وتُستَقَى مادَّتُها من مُختلف فُرُوع المعرِفة شريطَةَ أن تُبْرز واقِعَ اللُّغة المُعَيَّنة.

وإذا كانَ الأصلُ في شبكات الكلمات أن تكونَ قواعدَ بياناتٍ مُعجميَّة، تتلاحمُ عناصرُها باستعمال العلاقات الدَّلاليَّة، فإنَّها – أي شبكات الكلمات – تُمَثِّلُ مجموعةً من الأنطولوجيات الجُزئيَّة. وعلى هذا، فإنَّ وُجودَ شبكةٍ معيارِيَّةٍ للكلمات العربيَّة يفتحُ الطَّريقَ أمامَ بناء أنطولوجيا عربيَّة كاملة؛ وهو مشروعٌ طَموحٌ يعوقُ تحقيقَه ضخامةُ المُحتوى العربيّ وتنامِيه المُستمرّ، والغُموض في كثيرٍ من مفاهِيمِ هذا المُحتوى.

(5) Staab, S., Studer, R., (2004). Handbook on ontologies. Birkhäuser. And: Poli, R., Healy, M., Kameas, A., (2010). Theory and Applications of Ontology: Computer Applications. Springer.



الشَّكل ٣: العلاقات بين المفاهيم (الشَّبكة الدَّلاليَّة – شبكة الكلمات – الأنطولوجيا)

٢- الإشكاليَّة.

سارَت شبكة الكلمات العربيَّة AWN على منهج شبكة كلمات برينستون PWN في مادَّتها الأساسيَّة. وإذا أردنا أن نتحَرَّى الدِّقَّة، فيُمكن القول إنَّ شبكةَ الكلمات AWN ليسَت عربيَّةً بالمفهوم الدَّقيق، بل إنَّها ترجمةٌ لبعض أجزاء شبكة كلمات برينستون PWN.

وللوُقُوف على جوانب القُصُور الَّتي يمكن أن تنتج عن مثل هذه التَّرجمة، فثمَّةَ بعضُ الحقائق الَّتي تساعدنا على فهم طبيعة اللُّغات الإنسانيَّة وأوجه التَّبايُن بينها، نُحاولُ أن نُجمِلُها في النِّقاط التَّالية:

١. يستعمل الجنس البشريُّ ما يربو على سِتَّة آلاف وتسعمئة (٦٩.٩) لُغة

مُختَلِفة في الحديث والكِتابة، ولِكُلِّ لُغَةٍ من هذه اللُّغات مفرَداتُها الَّتي لها نظائِرُها في اللُّغات الأخرى[٦].

٢. البِيئة العَرَبيَّة هي الَّتي صَنَعت مِئات الكلمات لوصف الجَمَل والجَواد والسَّيف، وليس لذلك مَثيلٌ فيما عَداها مِن لُغات.

٣. في لُغة جرينلاند القُطبِيَّة ثَلاثٌ وسِتُّون كلمة بِمعنى (ثلج)[٧] .

٤. في لُغة هاواي مئة وثمانِي كلمات بِمعنى (بطاطا حُلوَة)[٨].

٥. لا وُجودَ للأرقام في قبائِل بيرو البِدائِيَّة إلاَّ (من 1 إلى 4)، ويُعَبَّر عن غير ذلك من الأرقام بِكلمة واحِدة هي (tūbāytē)[٩].

٦. تختلف دَلالَةُ المُفرَدات على المُذكَّر والمؤنَّث من لُغةٍ إلى أخرى، ونُلاحظُ هذا بين العربيَّة والألمانيَّة – مثلاً – في كلماتٍ مثل (قهوة Kaffee، طاولة Tisch) حيثُ تدُلُّ على مُؤَنَّثٍ في العربيَّة ومُذكَّرٍ في الألمانيَّة، ومثلها (الباب Tür) الَّتي تدُلُّ على مُذَكَّرٍ في العربيَّة ومؤنَّثٍ في الألمانيَّة.

إنَّ طبيعَةَ اللُّغات الإنسانيَّة Human Languages تقتضي أن تختلِفَ عناصِرُ الأنظمة اللُّغَوِيَّة وتَتَبايَنَ فيما بينها باختلافِ الزَّمان والمكان؛ وهذا حالُ اللُّغة العربيَّة، إذ

---

(6) Finegan, E., (2011). Language: its structure and use. Thomson Wadsworth. 6th Edition. and Ginsburgh, V., Weber, S., (2011). How many languages do we need?: the economics of linguistic diversity. Princeton University Press.

(٧) السَّعيد (المُعتزّ بالله): اللُّغة الإنسانيَّة، دار الهاني، القاهرة، ٢٠١١.

(٨) السَّعيد (المُعتزّ بالله): السَّابق.

(٩) السَّعيد (المُعتزّ بالله): السَّابق.

تتمتَّعُ ببعض الخصائص الَّتي تُمَيِّزُها عن غيرها من لُغات الفصائل اللُّغَوِيَّة الأخرى، بل وتُمَيِّزها – كذلك – عن بعض شقيقاتِها مِن اللُّغات السَّامِيَّة. فهي لُغَةٌ اشتِقاقيَّةٌ، يقوم نظامُها المُعجميّ على توليد الألفاظ والمعاني؛ وهي – كذلك – لُغةٌ مُعرَبةٌ، تتغيَّرُ أواخرُ كلماتها تبعًا لموقعها الإعرابيّ. وبعبارةٍ أخرى تتَّسم اللُّغة العربيَّة ببعض السِّمات الَّتي تُمَيِّزُها عن غيرِها من اللُّغات الإنسانيَّة على مُستوى المباني والمعاني.

فعلى مُستوى المباني، تتشكَّلُ ألفاظُ اللُّغة العربيَّة من أربعةٍ وثلاثين فونيمًا قطعيًّا Segmental Phoneme؛ وينتج عن التَّداخل بين الفونيمات (المنطوقة) تداخُلٌ بين الجرافيمات (المكتوبة)؛ وهذا ما يؤدِّي إلى تغيُّر شكل الحرف الواحد في الكلمة تبعًا لموضعه فيها من ناحية، وتعدُّد الألفاظ والمعاني المُنسدلة عن المجموعاتِ الكتابيَّةِ المُتَّفقةِ في حُرُوفها من ناحيةٍ أخرى.

الكلمة		PoS	المعنى
كتب			
كَتَبَ	Kataba	فِعل	Write
كَتْب	Katb	مَصدَر	Writing
كُتُب	Kotob	جَمْع	Books
بل			
بَلْ	Bal	أداة	But
بَلَّ	Balla	فِعل	Wet
بَلٌّ	Ball	اسْم	Drencher

الجدول ٢: نموذج – التَّداخُل (الفونيمي والجرافيمي) للأشكال الكتابيَّة في العربيَّة

أمَّا على مُستوى المعاني، فيبدو أثرُ البيئة العربِيَّة واضِحًا في تعدُّد دَلالة المُفرَدة الواحدة من ناحية، وتعدُّد المُفرَدات الَّتي تُشِيرُ إلى دَلالةٍ واحدةٍ من ناحيةٍ أخرى. فمثالُ تعدُّد دَلالة المُفرَدة الواحدة: كلمة **(حاجب)** الَّتي تُشِيرُ إلى العديد من الدَّلالات، منها:

الدَّلالة العربيَّة	المُقابِل الإنجليزِيّ
مُساعِد المحكمة	Usher
خازن الباب وحارِسه	Doorman
الشَّعْر النَّابِت فَوقَ العَيْن	Brow
عَلَم مُذَكَّر (من أسماء الذُّكور)	Hajeb

الجدول ٣: نموذج – تَعَدُّد دَلالة كلمة (حاجب) في العربيَّة

ومثالُ تعدُّد المُفرَدات الَّتي تُشِيرُ إلى دَلالةٍ واحدةٍ: الكلمات (سَبع / غَضَنفَر / حارِث / هِزَبر / أُسامة / حَيدَر / بَهنَس / لَيْث / قَسوَرة / فِرناس / باسِل / دِرغام / حَمزة / ...)، إذ تُشِيرُ جميعُها إلى (الأسَد) فيما يُعرَف بظاهِرة (التَّرادُف Synonym).

ووفقًا لمنهج الشبكة المُنجَزَة AWN، لا وُجودَ لهذه الكلمات في قاعدة البيانات المُعجَمِيَّة، إذ تُشِيرُ جميعُها إلى المُقابِل الإنكليزِيّ "Lion"، وإن كانت البيئة الإنكليزِيَّةُ، الَّتي وُجِدَ فيها الأسدُ مُنذُ القِدَم، تسمح بالتَّمييز بينَ دَلالات الكلمة، حيثُ يُشارُ إلى المُذَكَّر (أسد = Lion)، وإلى المُؤَنَّث (لَبُؤَة = Lioness) وإلى الصَّغِير (شِبل = Cub)؛ وهو أمرٌ لا نجدُه في كلمةٍ مُستَمَدَّةٍ من البيئة العربِيَّة، مثل كلمة

(جَمَل) الَّتي يُشارُ إلى دَلالاتِها الثَّلاث (المُذَكَّر «جَمَل»، والمُؤَنَّث «ناقة»، والصَّغير «حُوار») بالمُقابل الإنكليزيّ «Camel».

تكمُنُ إشكاليَّة التَّرجمة إذَن في أمرَين:

أ. اعتبار العربيَّةِ لُغةً لصقِيَّة، وهو ما يتنافى مع طبيعة مبانِيها.

ب. إغفال أثر البيئة في اللُّغة، وهُو ما يتنافى مع طبيعة معانِيها.

٣- المنهجِيَّة المُقترَحة.

في ضوء ما ذَكرناه حولَ مفهوم شبكة الكلمات وطبيعة الشَّبكات المُنجَزة، يُمكنُ القولُ إنَّه ليسَت لدينا – حتَّى الآن – شبكةٌ لكلمات اللُّغة العربيَّة بالمفهوم الدَّقيق لشبكات الكلمات؛ ولذا، تقترحُ الدِّراسةُ بناءَ شبكةٍ للكلمات العربيَّة انطلاقًا من المعايير المنهجيَّة الَّتي سارَت عليها شبكة كلمات برينستون PWN والشَّبكات اللُّغَوِيَّة للكلمات الَّتي أُنجِزَت تحتَ مِظَلَّة شبكة الكلمات العالميَّة Global WordNet، مثل: شبكة الكلمات الأوربيَّة Euro WordNet (EWN)[١٠]، وشبكة الكلمات الحُرَّة للُّغة الفرنسيَّة WordNet Libre du Français (WOLF)[١١]، والأنطولوجيا المِعيارِيَّة المُوَحَّدة – سومو Suggested Upper Merged Ontology (SUMO)[١٢]، وأنطولوجيا ياجو Yet Another Great Ontology (YAGO)[١٣]، وغيرِها.

ويُفترَض في مُحتوى شبكة الكلمات العربيَّة المنشودة أن يكونَ مناسبًا للمنطقِ اللُّغَوِيّ ومُوافقًا لطبيعة اللُّغة العربيَّة الاشتقاقيَّة ونظامها المُعجميّ، كما يُفترَضُ أن يُظهِرَ صُورةً للعربيَّة المُعاصرة على مُستوى المباني والمعاني، بما يتناسبُ مع الهدف من

(10) http://www.illc.uva.nl/EuroWordNet/.
(11) https://gforge.inria.fr/projects/wolf/.
(12) http://www.ontologyportal.org/.
(13) http://www.mpi-inf.mpg.de/yago-naga/yago/.

شبكات الكلمات في صناعة المُعجَمات اللُّغَوِيَّة وتطبيقات الويب الدَّلاليّ من ناحية، وبما يفتحُ الطَّريقَ أمامَ بناء الأنطولوجيا العربيَّة الكاملة من ناحيةٍ أخرى.

وإذا كانَ الأصلُ في شبكة الكلمات WordNet أن تكونَ قاعدةَ بياناتٍ مُعجَمِيَّة Lexical database، أو بعبارةٍ أخرى: أن تكونَ شبكةً من العلاقات اللُّغَوِيَّة المُتداخلة، لا تقتصر مادَّتُها على العلاقات الدَّلاليَّة بين المُفردات، وإنَّما تمتدُّ لتشملَ العلاقات النَّاتجة عن مستويات التَّحليل اللُّغَوِيّ الَّتي تتكوَّنُ عنها المُعجماتُ اللُّغَوِيَّة، فالدِّراسةُ تقترحُ ثلاثَ مراحل لبناء شبكةٍ لكلمات اللُّغة العربيَّة، هي (الإدخال «التَّعبئة»، والإثراء «الإغناء»، والرَّبطُ بين البيانات). وبيانُ هذه المراحل فيما يلي.

١.٣ – المرحلة ١: إدخال البيانات.

تعتمدُ هذه المرحلةُ على الموارد اللُّغَوِيَّة الأساسيَّة للشَّبكة المنشودة، وتتمثَّلُ في:

١.١.٣ – المُدَوَّنة اللُّغَوِيَّة Linguistic Corpus.

هي المصدر الأساسيّ لمُفردات شبكة الكلمات؛ وتُمَثِّلُ المُدَوَّنةُ اللُّغَوِيَّةُ – عمومًا – مجموعةً مِن نُصُوص اللُّغة الَّتي يمكن التَّعامل معها آلِيًّا والتَّحَكُّمُ في بياناتها ومُدخَلاتها – إضافةً أو حذفًا أو تعديلاً – بالاستفادة من قواعدِ بياناتٍ صُمِّمَت لتكونَ قادرةً على التَّعامُل مع هذه النُّصوص، حيثُ تُمَثِّلُ هذه القواعدُ مخزنًا كبيرًا للُّغة، يُرجَع إليه وقتَ الحاجة ويَتحمَّل أيَّ قدرٍ من النُّصوص الَّتي يُمكن أن تُضافَ إلى المادَّة الأساسيَّة للمُدَوَّنة اللُّغَوِيَّة مُستَقبلاً [١٤].

(١٤) السَّعيد (المُعتزّ بالله): مُدَوَّنة مُعجَم تاريخيّ للُّغة العربيَّة «مُعالَجة لُغَوِيَّة حاسوبيَّة»، رسالة دكتوراه، جامعة القاهرة، ٢٠١١م.

ويُفترَض في المُدَوَّنة اللُّغَوِيَّة المستعملة في تعبئة شبكة الكلمات العربيَّة أن تضُمَّ نُصُوصًا من العربيَّة الفُصحى المُعاصِرة، باعتبارِ هذا المُستوى (الفُصحى المُعاصِرة) تمثيلاً لواقع اللُّغة العربيَّة (على المستويين الجُغرافيّ والموضُوعيّ) من ناحية، ومُعَبِّرًا عن الاستعمال اللُّغَوِيّ للعربيَّة بينَ مُستَعمليها من ناحيةٍ أخرى.

٢.١.٣- مسرَد المُفرَدات / الألفاظ Glossary of Words.

المُفرَدة (أو الكلمة) – في اصطِلاح اللُّغَوِيِّين – صيغةٌ ذات وظيفة لُغَوِيَّة مُعَيَّنة في تركيب الجُملة، تقوم بِدَور وحدة من وحدات المُعجَم، وتصلح لأن تُفرَد أو تُحذَفَ أو تُحشى [بزيادة السَّوابق واللَّواحق والحشايا]، أو يُغَيَّر موضعها، أو يُستَبدَل بها غيرها في السِّياق، وترجع مادَّتها غالبًا إلى أصولٍ ثلاثة [١٣]. ويُعتَبَرُ مسرد المُفرَدات موردًا لُغَوِيًّا فرعيًّا، إذ يستمدُّ مداخلَه ووحداته من المُدَوَّنة اللُّغَوِيَّة بعد فهرسة نُصُوصِها آليًّا.

ولكن كيفَ تُصَنَّفُ المُفرَدات / الكلمات في شبكة الكلمات العربيَّة المنشودة؟!

لقد صَنَّفَ النُّحاةُ القُدَماءُ الكلامَ العَرَبِيَّ إلى مجموعةٍ من المُكَوِّنات / العَناصِر الَّتي تُعرَفُ بأقسام الكَلام PoS. وقَسَّمُوها إلى ثلاثَة أقسامٍ، هي: الاسم، والفِعل، والأداة (حرف لهُ معنى). وَيَحِيدُ بَعضُ اللُّغَوِيِّينَ المُحدَثِينَ عن هذا التَّقسِيم، حيثُ يَتَّجِه الدُّكتور إبراهِيم أنيس (١٩٠٦ – ١٩٧٧) إلى تَقسِيم أجزاء الكلام تَقسِيمًا رُباعِيًّا (الاسم، والفِعل، والحَرف، والضَّمِير)(١٥)، على حين يَتَّجِه الدُّكتور تمَّام حَسَّان (١٩١٨- ٢٠١١) وطائِفةٌ من الباحثِين إلى التَّقسِيم السُّباعِيّ (الاسم، والصِّفَة، والفِعل، والضَّمير، والخالِفة، والظَّرف، والأداة)(١٦).

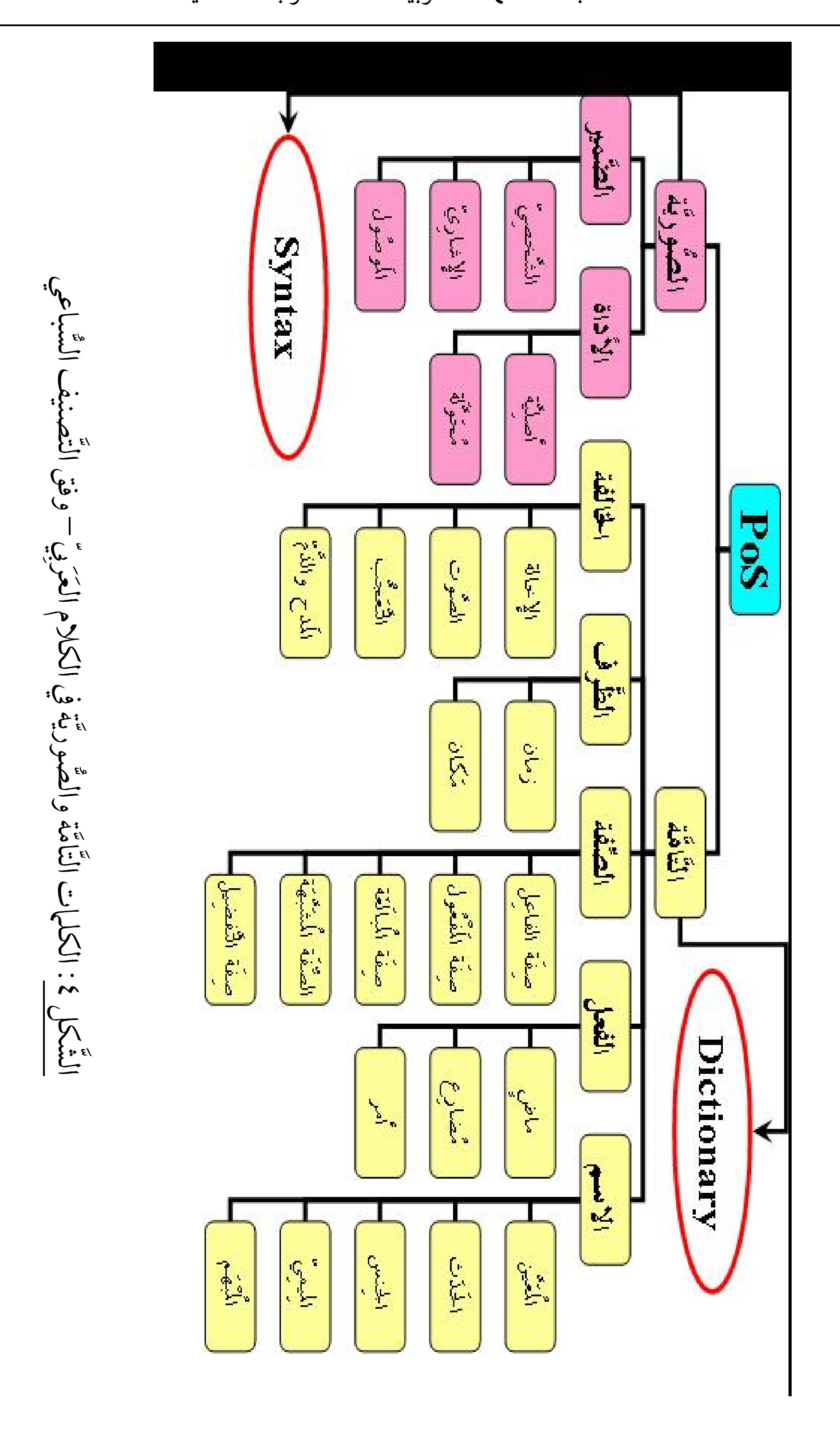
(١٥) أنيس (إبراهِيم): مِن أسرار اللُّغة، مكتبة الأنجلو المِصْريَّة، القاهِرة، ط٤، ١٩٧٢م.

(١٦) حَسَّان (تمَّام): اللُّغة العَرَبِيَّة "مَعناها وَمَبناها"، الهَيئة المِصرِيَّة العامَّة للكِتاب، القاهِرة، ط٢، ١٩٧٩م.

وتُصَنَّفُ الكلمات – عُمومًا – إلى كلماتٍ تامَّة Full words وكلماتٍ صُوريَّة Form words؛ فالكلمات التَّامَّة هي الَّتي تحمل معنًى تامًّا في ذاتها – ومصدرُها المُعجم (Dictionary)، وتُمَثِّلُها – وفق التَّقسيم السُّباعيّ: الأسماء (كتاب، قَلَم، مدرَسة، ...)، والأفعال (كَتَبَ، يفهم، يُغادِر، ...) والصِّفات (أخضَر، هادئ، سقيم،...)، والظُّرُوف (تحت، أسفل، خلف،...)، والخوالف (نِعمَ، بِئسَ، حبَّذا، ...)؛ أمَّا الكلمات الصُّوريَّة فهي الَّتي لا يُعرَف معناها إلاَّ بالنَّظر إلى الكلمات المُجاورة لها، أو الجُملة المرتبطة بها – ومصدرُها علم التَّركيب (Syntax)، وتُمَثِّلُها: الأدوات (مِنْ، يا، ما، هل، ...)، والضَّمائر (نحن، أنتَ، هي، هُم، ...).

وإذا كانَ الهدفُ الأساسيّ من بناء شبكة الكلمات العربيَّة أن تكونَ قاعدَةَ بياناتٍ مُعجَميَّة، وأداةً مُساعدةً في تقنيات الأنطولوجيا العربيَّة والويب الدَّلاليّ، فإنَّنا لن نكونَ بحاجةٍ إلى تزويد الشَّبكة المنشودة بقوائم الكلمات الصُّوريَّة (الأدوات والضَّمائر)، إذ الأَوْلى أن يُكتفى بالكلمات التَّامَّة الَّتي تحمل معنًى كاملاً يدلُّ عليها.

وإذا كانت شبكة كلمات برينستون PWN قد صنَّفت الكلامَ إلى أربعة أقسامٍ، هي: الاسم والفعل والصِّفة والظَّرف، وحاكَتها شبكةُ الكلمات العربيَّة AWN مُكتَفِيَةً بتقسيم الكلام إلى الاسم والفِعل، فإنَّ الدِّراسةَ تقترحُ تقسيمَ الكلام إلى خمسة أقسامٍ مُستَمَدَّةٍ من التَّقسيم السُّباعيّ العربيّ، هي: الاسم والفِعل والصِّفة والظَّرف والخالفة.



الشَّكل ٤: الكلمات التَّامَّة والصُّوريَّة في الكلام العَرَبِيّ – وفق التَّصنيف السُّباعي

٣.١.٣- مسرَد التَّعبيرات الاصطلاحيَّة Idioms Glossary.

تُشير الكلماتُ المُفرَدة إلى معانٍ مُعجَمِيَّة حرفيَّة أو مجازيَّة؛ وتتعدَّدُ هذه المعاني بِتَعَدُّد التَّنويعات السِّياقِيَّةِ وأساليب التَّعبِير؛ وعلى سَبِيل المِثال: تَتَعَدَّدُ المعانِي المُعجَمِيَّة لكلمة (عَيْن)؛ فقد تَعني «عُضو الإبصار»، وقد يُرادُ بِها «يَنبُوع الماء في الأرض»، وتَرمِي – مجازًا – إلى «الجاسوس»، وغيرِها مِن المعاني الَّتي تتَّضِحُ من السِّياق.

أمَّا التَّعبيرُ الاصطلاحيّ – الَّذي ينتُجُ عن تجمُّعٍ من الكلمات – فيُشيرُ إلى معنى مجازيٍّ ثابتٍ؛ ومن أمثلته في العربيَّة: التَّعبير (يغرق في شبر ماء) ومعناه (قليل الحيلة)، والتَّعبير (ضَرَبَ كفًّا بكفّ / ضرَبَ أخماسًا لأسداس) ومعناه (تَحيَّرَ).

وإثراءً لُمحتوى شبكة الكلمات العربيَّة المنشودة، تقترحُ الدِّراسةُ تزويدَها بمسرَدٍ للتَّعبيرات الاصطلاحيَّة المستعملة في العربيَّة المُعاصِرة، بحيثُ تُستَمَدُّ مادَّتُها من المُدَوَّنة اللُّغَوِيَّة. ويُفتَرَضُ أن يُمَثِّلَ هذا المسرَدُ موردًا لُغَوِيًّا فرعيًّا – على شاكلة مسرَد الكلمات.

٢.٣- المرحلة ٢: إثراء البيانات.

وتعتمدُ هذه المرحلةُ على الموارد اللُّغَوِيَّة الثَّانويَّة للشَّبكة المنشودة، وتتمثَّلُ في:

١.٢.٣- مُعجَم الألفاظ (Words Lexicon) Dictionary.

ويُمَثِّلُ مصدر المعاني المُعجَمِيَّة – الحَرفِيَّة والمجازِيَّة – في شبكة الكلمات المنشودة. وتقترح الدِّراسةُ جمعَ مادَّتِهِ من المُعجمات اللُّغَوِيَّة التَّالية:

- **مُعجَم العَرَبيَّة الحديثة** Arabisches Wörterbuch für die Schriftsprache der Gegenwart للمُستَشرِق الألمانِيّ هانز فِير Hans Wehr (1909–1981).
- **المُعجَم الوسيط**، الَّذي أخرَجه مجمع اللُّغة العربِيَّة بالقاهرة في طبعته الأولى عام ١٩٦٠ م، وشارك فى تَحرِيرِه نُخبَةٌ من المُعجَمِيِّين.
- **مُعجم اللُّغة العربيَّة المُعاصرة**، الَّذي حرَّرَه الدُّكتور أحمد مختار (١٩٣٣–٢٠٠٣) بمُساعدة فريق عمل، وظهرَت الطَّبعة الأولى منه في ٢٠٠٨ م.

٣.٢.٢– مُعجَم التَّعبيرات الاصطلاحيَّة Idioms Lexicon.

وهو مصدر المعاني المجازيَّة للتَّعبيرات الاصطلاحيَّة في الشَّبكة المنشودة. وتقترح الدِّراسةُ جمعَ مادَّتِهِ من المُعجَمَين التَّالي ذكرُهما:

- **قاموس المُتَعلِّم للتَّعابير الكلاسيكيَّة العربيَّة Learner's Dictionary of Classical Arabic Idioms، الَّذي حَرَّرَه الأستاذ ليزلي ماكلوكلين Leslie J. McLoughlin، ونُشِرَ في عام ١٩٨٨ م.**
- **المُعجم السِّياقيّ للتَّعبيرات الاصطلاحيَّة، الَّذي حرَّرَه الدُّكتور محمُود إسماعيل صيني وآخرون، ونُشِرَ في ١٩٩٦ م.**

٣.٣– المرحلة ٣: الرَّبط بين البيانات.

وتُعنى هذه المرحلة ببناء الشَّبكة العلائقِيَّة Relational Network بين مداخل الشَّبكة المنشودة. ويجري إنشاؤها في مرحَلَتَين فَرعِيَّتَين، نعرضُ لهما فيما يلي.

٣.٣.١– تعيين الحُقُول الدَّلاليَّة.

الحقلُ الدَّلاليُّ Semantic field هو مجموعة من الكلمات الَّتي ترتبط

دَلالاتها وتُوضَعُ تحتَ لفظٍ عامٍّ يجمعها[١٧] . وهناك العديد من التَّصنيفات الَّتي قُدِّمَت لتكونَ حُقُولاً دَلاليَّةً لُمفردات اللُّغة؛ لكنَّ أشملَ التَّصنيفات المطروحة وأكثرَها منطقيَّةً ومُلاءَمَةً لشبكة الكلمات العربيَّة المنشودة – من وجهة نظر الباحث – هو التَّصنيف الَّذي اقترَحه المُعجم اليونانيّ لُمفَردات العهد الجديد Greek New Testament، الَّذي حرَّره فريقٌ من اللُّغَوِيِّين، حيثُ يضُمُّ ما يقرُب من ٥٠٠٠ كلمة، تحوي قوائِمُها ١٥٠٠٠ معنى مُعجَمِيّ؛ وتندرِجُ فيه المُفردات في أربعة حُقُولٍ دَلاليَّة[١٨]، هي: الموجودات Entities، والأحداث Events، والمُجرَّدات Abstracts، والعلاقات Relations. وتتفَرَّعُ هذه الحُقُول الرَّئيسية لتشمل حُقُولاً فرعيَّة، تتفرَّعُ بدورها إلى حُقُولٍ فرعيَّةٍ أخرى، وهكذا.

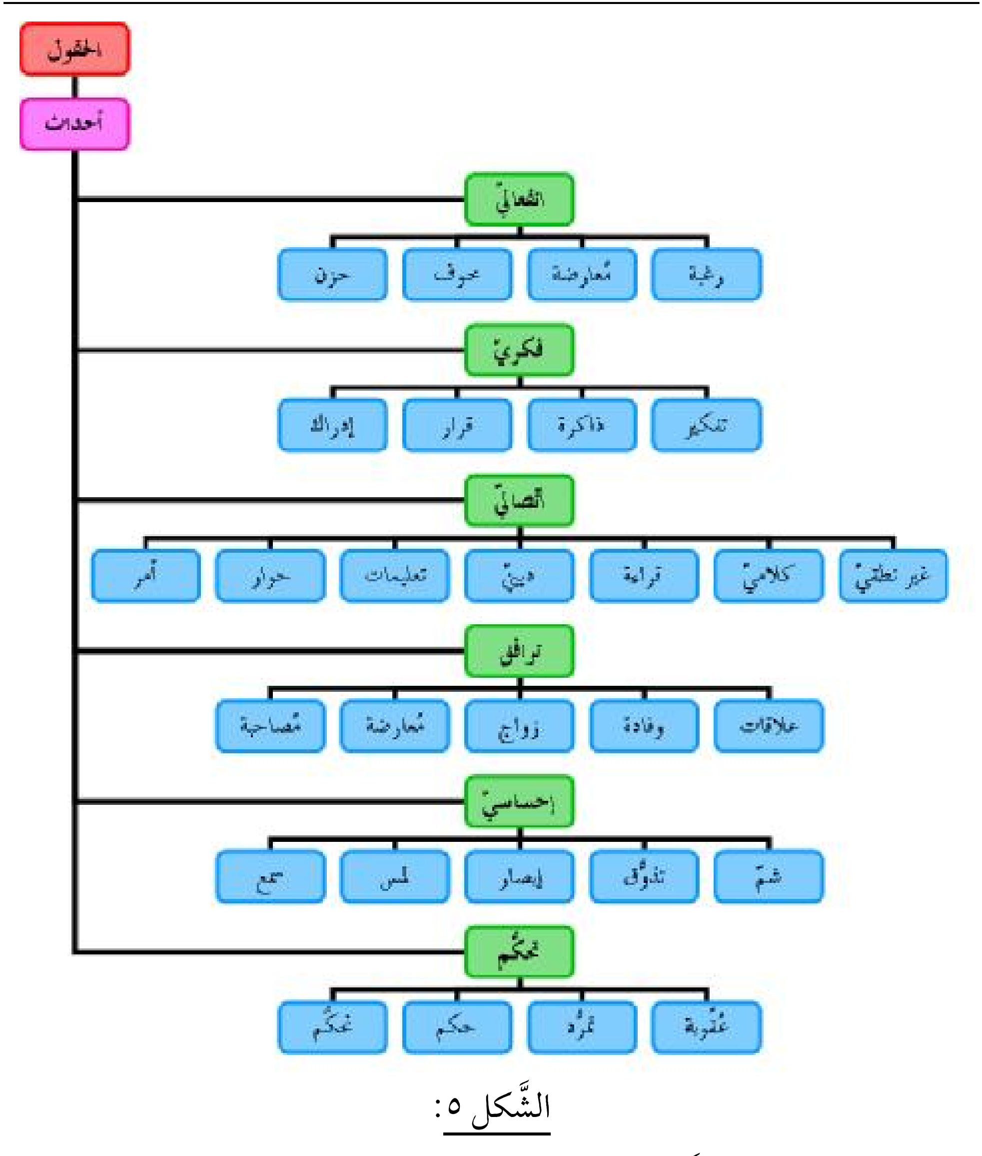
وتقترح الدِّراسةُ – إضافةً إلى هذا المُعجم – الاستعانة بالمصادر التَّالية:

- كتاب «مفهوم **النِّظام** بوصفه أساساً لصناعة المُعجم: مُحاولة لتصنيف النِّظام» Begriffssystem als Grundlage für die Lexikographie: Versuch eines Ordnungsschemas، الَّذي حرَّرَه الألمانيَّان رودولف هالينج Rudolf Hallig وفالتر فون فارتبورج Walther von Wartburg، ونُشِرَ في ١٩٦٣ م.
- مُعجم «المُخَصَّص»، لابن سِيده الأندلسيّ (ت ٤٥٨ هـ).
- مُعجم «المُنَجَّد في اللُّغة»، لكُراع النَّمل الهُنائيّ (ت ٣١٠ هـ).

---

(١٧) عُمَر (أحمد مُختار): علم الدَّلالة، عالم الكُتُب، القاهرة، ط ٥، ١٩٩٨ م.

(18) United Bible Societies: The Greek New Testament. ISBN 3-438-05110-9.

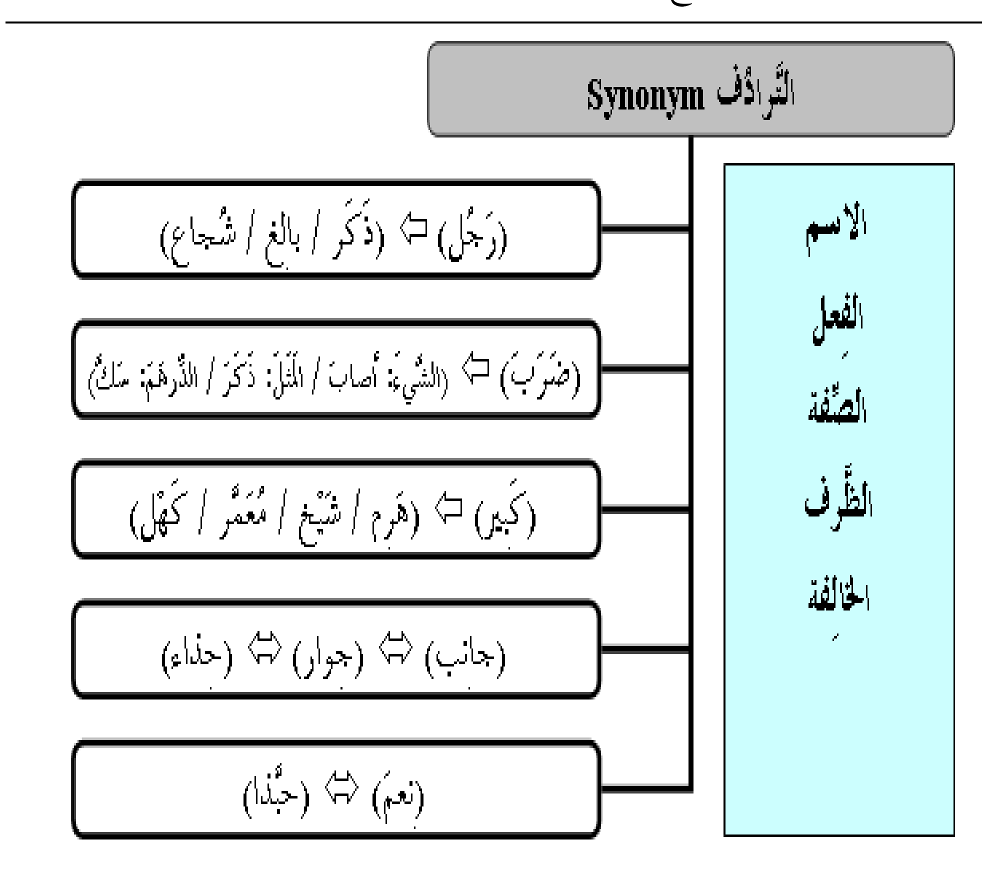


الشَّكل ٥:

نموذج من الحقل الدَّلاليّ (الأحداث) – عن مُعجم Greek New Testament

٣.٣.٢ – تعيين العلاقات الدَّلاليَّة.

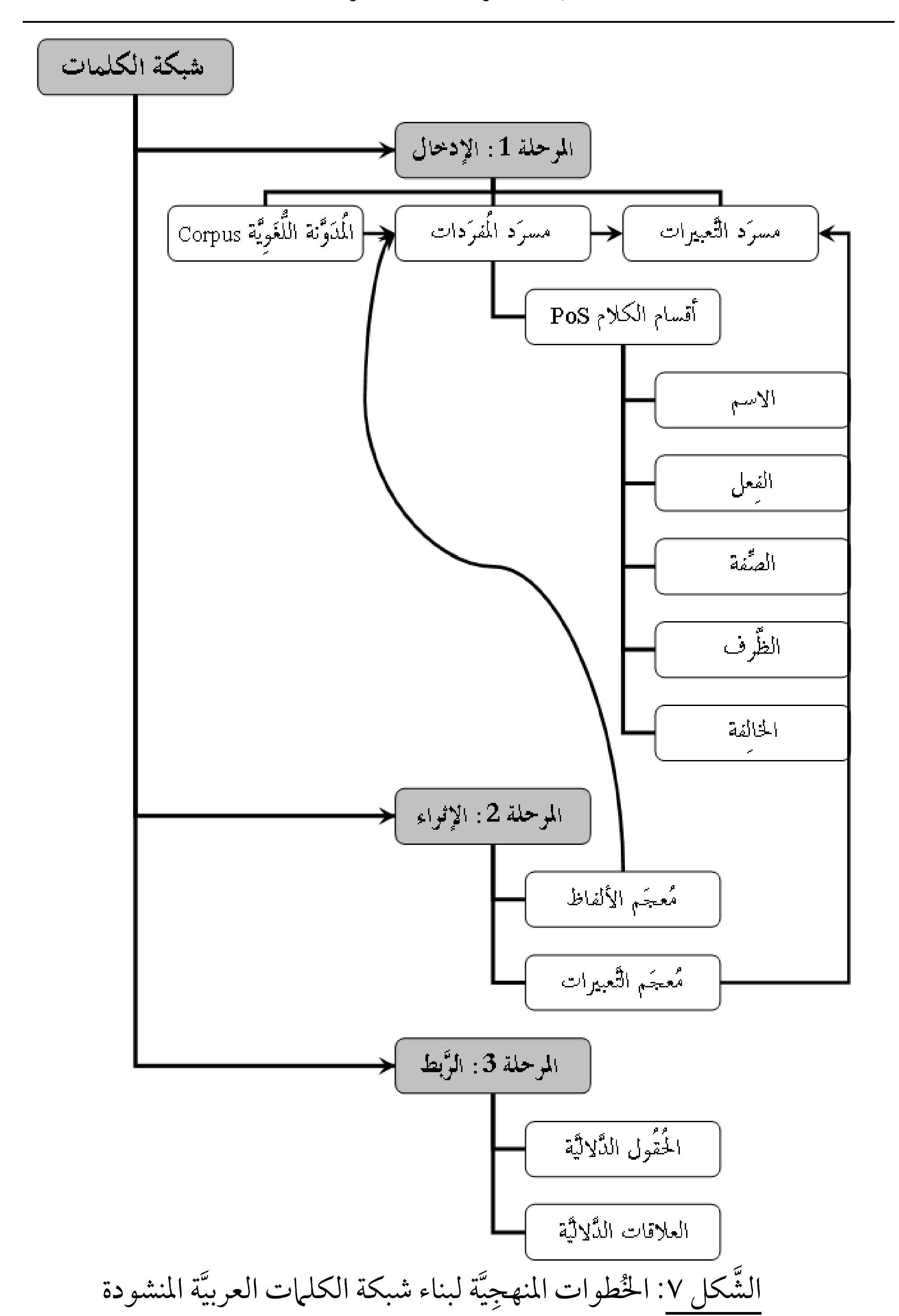
تقترح الدِّراسةُ الإفادة من العلاقات الدَّلاليَّة المُدرجة في شبكة كلمات برينستون PWN بتطبيقها على مُفردات العربيَّة – وفقَ التَّقسيم الخُماسيّ، على النَّحو المُبيَّن في المثال التَّالي (مثال من العربيَّة لعلاقة التَّرادُف في أقسام الكلام الخمسة).



الشَّكل ٦: علاقة التَّرادُف في أقسام الكلام العربيّ

وثمَّةَ العديدُ من المصادر الَّتي تُعنى بالعلاقات الدَّلاليَّة بين مُفردات اللُّغة العربيَّة، والَّتي يُمكن الإفادةُ منها في تعيين العلاقات الدَّلاليَّة بين مُفرَدات شبكة الكلمات المنشودة؛ وتقتَرِحُ الدِّراسةُ مصدرَين رَئِيسَيين – إضافةً إلى شبكة برينستون PWN، هما:

- المكنَز الكبير، الَّذي حرَّرَه الدُّكتور أحمد مختار عُمَر بمُساعدة فريق عمل، ونُشِرَ في عام ٢٠٠٠ م.
- المَكنَز العربيّ المُعاصِر، الَّذي حرَّرَه الدُّكتور محمُود إسماعيل صيني وآخرون، ونُشِرَ في ١٩٩٣ م.



الشَّكل ٧: الخُطوات المنهجِيَّة لبناء شبكة الكلمات العربيَّة المنشودة

٤ – التَّطبيق.

ثمَّةَ العديدُ من التَّطبيقات الَّتي تُمثِّلُ جوانبَ إفادةٍ من شبكة الكلمات العربيّة المنشودة، والَّتي ترتبطُ بآليَّة التَّحليل الدَّلاليّ للعربيَّة Semantic analysis. من هذه التَّطبيقات: تنمية المُحتوى الإلكترونيّ للعربيَّة على الشَّبكة العنكبوتيَّة Internet – سواءٌ أكانَ مُحتوى حُرًّا On-Line أم مُخَزَّنًا Off-Line، وتطوير عمل مُحرِّكات البحث العربيَّة Arabic Search engines، والمُساهمة في بناء وإثراء الأنطولوجيا العربيَّة المنشودة Arabic Ontology، وتنمية بيئة الويب الدَّلاليّ، وتطبيقات صناعة المُعجَم العربيّ؛ والتَّطبيقُ الأخيرُ هو الَّذي يعنينا في هذه الدِّراسة.

إنَّ الاعتمادَ على الموارد اللُّغَوِيَّة (المُدَوَّنة اللُّغَوِيَّة، ومُعجم المُفرَدات / الألفاظ، ومُعجم التَّعبيرات الاصطلاحيَّة) يُثمِرُ هيكلاً مُعجَمِيًّا يتوافقُ مع مُعجمات الألفاظ؛ كما يُثمرُ الرَّبطُ بين الموارد اللُّغَوِيَّة باستعمال الحُقُول الدَّلاليَّة – من ناحية – والعلاقات الدَّلاليَّة – من ناحيةٍ أخرى –هيكلاً مُعجَمِيًّا يتوافقُ مع مُعجَمات الموضوعات.

ويُمكنُ القولُ بعبارةٍ أخرى: إنَّ الالتزامَ بمعايير بناء شبكات الكلمات – مع مُراعاة مُناسَبَتِها لطبيعة اللُّغة العربيَّة على مُستوى المباني والمعاني – عندَ إعداد شبكة الكلمات العربيَّة المنشودة سيُثمرُ قاعدةَ بياناتٍ مُعجَمِيَّةٍ هائلةً تتشكَّلُ في هيكلٍ مُعجَمِيٍّ ضخمٍ، يجمعُ بين مُعجمات الألفاظ ومُعجمات الموضوعات، ويصلُح موردًا لُغَوِيًّا للعديد من التَّطبيقات المُتَعلِّقة بقضايا اللُّغة العربيَّة عُمومًا وتقنياتها على وجه الخُصُوص.

ومن هذا الهيكل المُعجَميّ الَّذي يتشَكَّلُ في مجموعةٍ من الحُقُول / المجالات الدَّلاليَّة، نستطيعُ بناءَ العديد من المُعجمات اللُّغَوِيَّة المُتَخَصِّصة، الَّتي يدعو إليها تنامي المعارف ونُشُوءُ العديد من العُلُوم البَينِيَّة (كحوسبة اللُّغة، والهندسة الطِّبِّيَّة، والرِّياضيَّات الفيزيقِيَّة)؛ وهذا ممَّا يدعمُ تعريبَ العُلوم ويُساعد على إثراء المُحتوى العَربيّ.

٥- نتائج الدِّراسة.

وبعد... فقد خلَصَت الدِّراسة إلى النَّتائج التَّالية:

١. رأت الدِّراسةُ أنَّ شبكة الكلمات العربيَّة المُنجَزة AWN ليسَت شبكةً للكلمات بالمفهوم الدَّقِيق، وإنَّما هي ترجمةٌ أو مُحاكاة لبعض أجزاء الشَّبكة الإنجليزِيَّة «برينستون» PWN؛ وقد أدَّت هذه التَّرجمةُ إلى قُصُورٍ منهجِيّ، تمثَّلَ في: اعتبار العربيَّةِ لُغةً لصقِيَّة، وإغفال أثر البيئة في اللُّغة.

٢. اقترَحَت الدِّراسةُ منهجِيَّةً لبناء شبكةٍ للكلمات العربيَّة انطلاقًا من شبكة كلمات برينستون PWN، مع مُراعاة طبيعة اللُّغَة العربيَّة. وتقوم المنهجِيَّة المُقتَرَحة على ثلاث مراحل أساسيَّة، هي (الإدخال، والإثراء، والرَّبطُ بين البيانات).

٣. اقترَحَت الدِّراسة تقسيم الكلام في الشَّبكة المنشودة إلى خمسة أقسام، هي: الاسم، والفعل، والصِّفة، والظَّرف، والخالفة.

٤. اقترَحَت الدِّراسةُ الإفادة من مواردَ لُغَوِيَّة مُعَيَّنة في تعبئة شبكة الكلمات المنشودة وإثراء مُحتواها، وتَمَثَّلَت في: (المُدَوَّنة اللُّغَوِيَّة، ومَسرَدَي المُفرَدات والتَّعبيرات، ومُعجَمَي المُفرَدات والتَّعبيرات، ومُعجَمَي الحُقُول الدَّلاليَّة والعلاقات الدَّلاليَّة).

٥. عَرَضَت الدِّراسةُ لبعض تطبيقات شبكة الكلمات، لا سيَّما في صناعة المُعجَم.

٦- الخلاصة

سَعَت هذه الورقةُ لِلتَّمييز بين المُصطَلحات الثَّلاثة (شبكة الكلمات، والشَّبكة الدَّلاليَّة، والأنطولوجيا)؛ وعَرَضَت الدِّراسةُ لإشكاليَّة شبكة الكلمات العربيَّة، كما قدَّمَت منهجًا مُقتَرَحًا لبناء وإثراء الشَّبكة المنشودة بما يتوافَقُ مع طبيعة اللُّغة العربيَّة، وبما يتناسَبُ مع تطبيقات الشَّبكة في تقنيات اللُّغة العربيَّة وصِناعة المُعجَم العَرَبِيّ.



٧- المصادر والمراجع

(1) Fellbaum, C., (1998). WordNet: an electronic lexical database. Cambridge, Mass: MIT Press

(2) Black, W., Elkateb, S., Rodriguez, H, Alkhalifa, M., Vossen, P., Pease, A. and Fellbaum, C., (2006). Introducing the Arabic WordNet Project, in Proceedings of the Third International WordNet Conference, Sojka, Choi, Fellbaum and Vossen eds.

(3) Staab, S., Studer, R., (2004). Handbook on ontologies. Birkhäuser.

(4) Poli, R., Healy, M., Kameas, A., (2010). Theory and Applications of Ontology: Computer Applications. Springer.

(5) Finegan, E., (2011). Language: its structure and use. Thomson Wadsworth. 6th Edition.

(6) Ginsburgh, V., Weber, S., (2011). How many languages do we need?: the economics of linguistic diversity. Princeton University Press.

(٧) السَّعيد (المُعتَزّ بالله): اللُّغة الإنسانِيَّة، دار الهاني للنَّشر، القاهرة، ط١، ٢٠١١م.

(8) http://www.illc.uva.nl/EuroWordNet/

(9) https://gforge.inria.fr/projects/wolf/

(10) http://www.ontologyportal.org/

(11) http://www.mpi-inf.mpg.de/yago-naga/yago/

(١٢) السَّعيد (المُعتزّ بالله): مُدَوَّنة مُعجَم تاريخيّ للُّغة العربيَّة «مُعالجَة لُغَوِيَّة حاسوبِيَّة»، رسالة دكتوراه، جامعة القاهرة، ٢٠١١م.

(١٣) حَسَّان (تمَّام): مناهج البحث في اللُّغة، مكتَبَة الأنجلو، القاهِرة، ط١، ١٩٥٥م.

(١٤) أنيس (إبراهِيم): مِن أسرار اللُّغة، مكتبة الأنجلو المِصْريَّة، القاهِرة، ط٤، ١٩٧٢م.

(١٥) حَسَّان (تَمَّام): اللُّغة العَرَبِيَّة «مَعناها وَمَبناها»، الهَيئة المِصرِيَّة العامَّة للكِتاب، القاهِرة، ط٢، ١٩٧٩م.

(١٦) عُمَر (أحمد مُختار): علم الدَّلالة، عالم الكُتُب، القاهرة، ط٥، ١٩٩٨م.

(17) United Bible Societies: The Greek New Testament.
ISBN 3-438-05110-9.

